09/819, 252 AA

Int. J. Cancer (Pred. Oncol.): **74**, 35–44 (1997)

Publication of the International Union Against Cancer
Publication de l'Union Internationale Contre le Cancer

MOLECULAR CLONING, SEQUENCING AND EXPRESSION OF THE mRNA ENCODING HUMAN Cdx1 AND Cdx2 HOMEOBOX. DOWN-REGULATION OF Cdx1 AND Cdx2 mRNA EXPRESSION DURING COLORECTAL CARCINOGENESIS

Gustavo V. MALLO,[1] Hocine RECHRECHE,[1] Jean-Marc FRIGERIO,[1] Dominique ROCHA,[2] Alain ZWEIBAUM,[3] Michel LACASA,[3] Bertrand R. JORDAN,[2] Nelson J. DUSETTI,[1] Jean-Charles DAGORN[1] and Juan L. IOVANNA[1]*

[1]*U.315 INSERM, Marseille, France*
[2]*Centre d'Immunologie INSERM-CNRS de Marseille-Luminy, France*
[3]*U.178 INSERM, Villejuif, France*

Defining the molecular mechanisms involved in cancer formation and progression is still a major challenge in colorectal-cancer research. Our strategy was to characterize genes whose expression is altered during colorectal carcinogenesis. To this end, the phenotype of a colorectal tumour was previously established by partial sequencing of a large number of its transcripts and the genes of interest were selected by differential screening on high-density filters with mRNA of colorectal cancer and normal adjacent mucosa. Fifty-one clones were found over-expressed and 23 were under-expressed in the colorectal-cancer tissues of the 5 analyzed patients. Among the latter, clones 6G2 and 32D6 were found of particular interest, since they had significant homology with several homeodomain-containing genes. The highest degree of similarity was with the murine Cdx1 for 6G2, and with the murine Cdx2 and hamster Cdx3 for 32D6. Using a RT-PCR approach, complete sequence of both types of homeobox-containing cDNA was obtained. The amino-acid sequence of the human Cdx1 is 85% identical to the mouse protein, and human Cdx2 has 94% identity with the mouse Cdx2 and hamster Cdx3. Tissue-distribution analysis of Cdx1 and Cdx2 mRNA showed that both transcripts were specifically expressed in small intestine, in colon and rectum. Twelve tissue samples from colorectal adenocarcinomas and the corresponding normal mucosa were analyzed by Northern blot. Expression of the 2 types of mRNA was either reduced or absent in 10 of them. Several colon-cancer cell lines were also analyzed. Cdx2 mRNA was absent from LS174T cells and Cdx1 mRNA was absent in PF11, TC7 and SW480 cells; none was detected in HT29 cells. It was concluded that decrease in human Cdx1 and/or Cdx2 expression is associated with colorectal tumorigenesis. *Int. J. Cancer 74:35–44.*
© 1997 Wiley-Liss, Inc.

Colorectal cancer is the second most common tumour in men and the third in women in the Western countries. Although much is known about the epidemiology, morphology and genetics of colorectal tumorigenesis, our knowledge of the perturbations of gene expression that occur in colorectal tumours remains limited. Such tumours may arise from benign adenomatous polyps, which later progress to adenocarcinomas through several molecular events. They thus provide a very useful paradigm for studying the molecular genetic bases of cancer. The multistep process leading to colorectal tumorigenesis probably involves the loss of function of tumour-suppressor genes, as well as the activation of oncogenes. Several important genes have already been identified, but they do not account for the whole process, and other genes are probably involved.

Efforts have been made to characterize these genes. Differential hybridization techniques and screening of substracted libraries allowed the elucidation of some of them (Bartsch *et al.*, 1986; Denis *et al.*, 1993; Yow *et al.*, 1988; Schweinfest *et al.*, 1993; Kondoh *et al.*, 1992; Barnard *et al.*, 1992*a*, *b*). We developed an alternative strategy, in which the phenotype of a colorectal tumour was established by partial sequencing of a large number of randomly selected transcripts (Frigerio *et al.*, 1995). This repertory of ESTs should therefore contain most of the differentially expressed genes. Recently, Nguyen *et al.* (1995) have developed an efficient method of differential screening in which cDNA clones are gridded on high-density colony filters and hybridized with complex probes derived from poly (A)$^+$ RNA from different cells or tissues. The signals observed are measured, providing a "hybridization signature" that characterizes the amount of mRNA produced by the corresponding gene in the tissue from which the probe was derived. cDNA corresponding to sequences present at levels of approximately 0.01% in the complex probes can be detected.

We have employed this method to select from our ESTs repertory genes differentially expressed in the colorectal cancer. This approach allowed us to select several clones that were over-expressed or under-expressed during colorectal carcinogenesis. We report here the cloning of the human Cdx1 and Cdx2 homeobox mRNA which are concomitantly down-regulated in about 85% of the tumours. Cdx1 and Cdx2 were of particular interest, since intestine was the only tissue among several adult organs surveyed which contained detectable levels for these transcripts.

MATERIAL AND METHODS

cDNA library and high-density filters

A cDNA library was constructed in pT7T3D (Pharmacia, Les Ulis, France) from poly(A)$^+$ RNA of an adenocarcinoma of the left colon with ganglionar and liver metastases obtained from an adult male. The unamplified library contained a total of 2.09×10^5 independent recombinant clones. Fresh transformant colonies were picked, grown in microwell plates and stored at −80°C in the presence of 15% glycerol. From this cDNA library, 2166 randomly selected clones were partially sequenced. In a first step, after exclusion of 27 clones corresponding only to the vector or poly(A) sequences, database search revealed that 1248 of our clones were already well characterized, 142 clones corresponded to known ESTs, while 891 clones (representing 740 novel genes) were absent from the database. These novel sequences have been deposited in the GenBank database with accession numbers T24426 to T25165. Clones encoding novel genes and ESTs were picked into 96-well plates. After picking, the bacteria were grown overnight in LB medium containing 15% glycerol and stored at −80°C.

High-density colony filters were prepared using a BIOMEK 1000 (Beckman, San Ramon, CA) robotic workstation and a 96-pin

Contract grant sponsor: Groupement de Recherches et d'Etudes sur les Génomes (GREG); Contract grant sponsor: Ligue Nationale Contre le Cancer (Comité des Bouches du Rhône); Contract grant sponsor: Association pour la Recherche sur le Cancer (ARC); Contract grant sponsor: Fédération Nationale des Centres de Lutte Contre le Cancer (FLNCC).

*Correspondence to: Juan L. Iovanna, INSERM U.315, 46 Bd de la Gaye, F-13009 Marseille, France. E-mail: iovanna@infobiogen.fr

Received 28 June 1996; Revised 11 October 1996

tool. Colonies from freshly grown replica plates were spotted onto Hybond-N filters (Amersham, Les Ulis, France) in a regular array as described by Nguyen *et al.* (1995). Filters were subsequently treated as described by Nizetic *et al.* (1991), with one modification, *i.e.*, increase in the duration of the proteinase-K treatment of the filters to 2 hr at 37°C (250 µg/ml) to achieve more complete removal of proteins.

Preparation and labelling of complex probes and hybridization conditions

RNA was isolated by the guanidium-thiocyanate method (Chirgwin *et al.*, 1979) from colorectal adenocarcinomas and the corresponding normal mucosa of 5 patients. Before labelling, 1 µg of poly(A)$^+$ RNA was heated to 65°C and then cooled to remove secondary structure in the RNA. Complex probes were then prepared by simultaneous reverse transcription and labelling in the presence of 100 µCi of [^{32}P]dCTP, using an oligo $dT_{(12-18)}$ as primer.

Filters were pre-hybridized for 4 hr at 42°C in a buffer containing 50% de-ionized formamide, 2% SDS, 4× SSC (1× SSC is 150 mM NaCl, 15 mM sodium citrate), 50 mM NaPi, 8% dextran sulphate, 0.2% BSA, 0.2% Ficoll, 0.2% polyvinylpyrrolidone and 1 mM EDTA. Hybridization was conducted at 42°C for 16 hr in the same buffer in the presence of the probe derived from poly(A)$^+$ RNA of a colorectal adenocarcinoma labelled as above. Filters were used in duplicate or triplicate. After hybridization, filters were washed once in 2× SSC at room temperature, then successively in 2× SSC, 0.1% SDS, 0.5× SSC and 0.2× SSC, all at 65°C for 20 min. Moist filters were then wrapped in plastic and exposed to X-ray film or to a Fuji (Tokyo, Japan) imaging plate. Filters were stripped as recommended by the manufacturer (Amersham), and then hybridized with the complex probe synthesized with the poly(A)$^+$ RNA of the corresponding normal mucosa and exposed to the Fuji imaging plate. The same procedure was applied for the complex probes derived from samples of 5 patients.

Quantitation of the hybridization signals

Autoradiography on X-ray film was used only for orientation or display purposes. Quantitation data were obtained using an imaging plate device. The hybridization filter was exposed to an image plate for 16 to 24 hr and then scanned in a FUJIX BAS 1000 (Fuji) system. These results represented the distribution of radioactivity on the filters, with linear response over 4 orders of magnitude. Hybridization signatures were determined by a modified version of the Bioimage software (Millipore, Bedford, MA) running on a Unix workstation. The resulting array of quantified spots was then matched with a template, a reference image that contained the complete grid of 1536 spots. This step eliminated most of the artefacts and assigned the correct name to each colony. This process generated a table containing the intensity and surface of each spot. The file was then exported to an Excell worksheet on a microcomputer and data were analyzed.

Isolation of cDNA encoding human Cdx1 and Cdx2

Using this approach, 234 of the 882 cDNA clones were detected as differentially expressed in colorectal carcinoma with the complex probes derived from the first patient. A similar proportion of differentially expressed ESTs was obtained with each of the 4 other patients. However, only 23 clones were constantly down-regulated and 51 over-expressed in the 5 tumours. We therefore completed the sequence of the insert of all these 74 cDNA clones. These sequences were then compared with a new release of the GenBank database. Among these genes, clones 6G2 and 32D6, which were down-regulated in colorectal carcinoma, had significant homology with several genes containing a homeodomain. The highest degree of similarity was with the murine Cdx1 (Duprey *et al.*, 1988; Hu *et al.*, 1993) and Cdx2 (James *et al.*, 1994; Suh *et al.*, 1994) mRNA, homologues of the *Drosophila* homeobox gene caudal (cad).

Sequencing of human Cdx1 and Cdx2 cDNA

Clone 6G2, which contains nucleotides from position 1065 to 1711 of the Cdx1 cDNA, and clone 32D6, which contains nucleotides from position 1007 to 1751 of the Cdx2 cDNA, were completely sequenced using the Sequenase version 2.0 DNA sequencing kit (USB, Cleveland, OH) and appropriate synthetic oligonucleotides. In order to complete the sequence of both Cdx1 and Cdx2 mRNA, we screened about 5 × 10^5 clones from a random primed cDNA library in λgt-11 made with RNA of a human normal colon excised around a colon cancer (Clontech, Palo Alto, CA) as described (Dusetti *et al.*, 1993) and using cDNA inserts of the clones 6G2 or 32D6 as probes. No clone could be found when the 6G2 was used as probe, and only one clone (λcdx2a) was found with the 32D6 probe. The λcdx2a cDNA insert was purified, sub-cloned into the pBluescript KS$^+$ and completely sequenced on both strands. Clone λcdx2a contained sequence from position 624 to 1110 of the Cdx2 mRNA. When the cDNA insert of the λcdx2a clone was used as probe to screen about 5 × 10^5 clones of the same cDNA library only 2 clones were selected. Complete sequence of both clones showed that they were identical to the λcdx2a insert.

Then, to clone the 5′-end of Cdx1 and Cdx2 mRNA, we developed the following RT-PCR-based strategy. Primers in 3′, designed from the sequence that we had obtained, were 5′GGGAATTCGGGCTTGAGATGCCCACGGGAG3′ for Cdx1 and 5′GGGAATTCTTTCCTCTCCTTTGCTCTGCGG3′ for Cdx2. Primer sequences in 5′ were based on the published murine Cdx1 mRNA sequence (Duprey *et al.*, 1988; Hu *et al.*, 1993) and murine (James *et al.*, 1994; Suh *et al.*, 1994) and hamster (German *et al.*, 1992) Cdx2 and Cdx3 mRNA sequences respectively. Cdx1 was specifically amplified with the 5′GGGAATTCAGGTGAGCAGTCGCTGGTCGTC3′ primer and Cdx2 with the 5′GGGAATTCGCGCC(C/T)CTGGCAGCCT(C/T)CA(A/G)CG3′ primer. Total RNA was extracted from a normal colon mucosa by the guanidinium-thiocyanate method (Chirgwin *et al.*, 1979). Oligo (dT)-primed cDNA was synthesized from total RNA (1 µg) using a cDNA synthesis kit (Boehringer Mannheim, Mannheim, Germany) following the manufacturer's instructions. The reaction mixture was heated at 70°C for 10 min and phenol/chloroform extracted. cDNA was PCR-amplified in 1× PCR buffer (50 mM KCl, 10 mM Tris HCl, pH 8.3, 2 mM $MgCl_2$, 0.01% gelatine) containing 125 µM dNTP, 25 pmol of each primer, and 2.5 units of Taq polymerase in a final volume of 50 µl. Amplification was performed according to "touchdown" protocol described by Don *et al.* (1991).

Tissue specimens and Northern blots

Normal tissues and tissue samples from colorectal adenocarcinomas and the corresponding normal mucosa were obtained (n = 12) from the Service de Chirurgie Digestive, Hôpital Sainte Marguerite (Marseille, France), in accordance with institutional review-board standards and procedures. Human colocarcinomas were classified and graded according to WHO: Dukes A, patients 4, 6, 7, and 12; Dukes B, patients 5 and 8; Dukes C, patients 9, 10 and 11; Dukes D, patients 1, 2 and 3. LS174T, PF11, TC7, SW480 and HT29 cell lines were grown as described (Chantret *et al.*, 1988). Total RNA was extracted from normal mucosa and tumours using guanidium thiocyanate (Chirgwin *et al.*, 1979), and RNA from the cell lines was purified using the acidic guanidinium thiocyanate-phenol-chloroform procedure (Chomczynski and Sacchi, 1987). For Northern-blot hybridization, 20 µg of denatured RNA were loaded onto each gel lane and submitted to electrophoresis on a 1% agarose-formaldehyde gel (Sambrook *et al.*, 1989) and transferred onto nylon filters (Hybond, Amersham). Filters were pre-hybridized for 4 hr at 65°C in a buffer containing 5× SSPE (1× SSPE is 180 mM NaCl, 1 mM EDTA, 10 mM NaH_2PO_4, pH 7.5), 0.1% BSA, 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.5% SDS and 200 µg/ml denatured herring sperm DNA. Hybridization was conducted at 65°C for 16 hr in the same buffer in the presence of the ^{32}P-labelled

6G2 and 32D6 cDNA inserts as probe. The filters were then washed 4 times for 5 min at room temperature in 2× SSC, 0.1% SDS, twice for 15 min at 65°C in 0.1× SSC, 01% SDS, and once for 30 min in 0.1× SSC. Filters were dried and exposed to Hyperfilm (Amersham).

RESULTS

Isolation of Cdx1 and Cdx2 cDNA

To select genes differentially expressed during colorectal carcinogenesis from our ESTs library (Frigerio *et al.*, 1995), we used a differential screening approach on high-density filters, as described by Nguyen *et al.* (1995). Filters were hybridized successively with complex probes prepared from poly(A)$^+$ RNA of colorectal cancer and normal adjacent mucosa. Clones 6G2 and 32D6 were selected because both are down-regulated in colorectal-cancer tissues. To complete their cDNA sequence, we have screened a cDNA library made with mRNA from a normal colon. Only one clone (λcdx2a) was isolated with the 32D6 cDNA insert as probe and none with the 6G2 cDNA insert as probe. To determine the 5′ end sequence of the Cdx1 and Cdx2 mRNA, we decided to use a RT-PCR approach with specific primers at the 3′. The 5′ primers were designed based on murine Cdx1 and murine Cdx2 and hamster Cdx3 respectively. After cloning the PCR products, sequencing was performed on both strands.

Cdx1 sequence: concatenation of the sequences yielded a mRNA sequence of 1695 nucleotides, exclusive of the poly(A) tail. The sequence of the Cdx1 is shown in Figure 1. A single open reading frame of 265 amino-acids was found in the cDNA sequence. The proposed ATG start codon at position 82 of the Cdx1 mRNA conforms to the Kozak consensus sequence and is in accordance with the mouse Cdx1 mRNA (Duprey *et al.*, 1988; Hu *et al.*, 1993). After the translational stop codon at position 877, a long 818-bp 3′ untranslated sequence follows, with a polyadenylation signal at position 1675.

Cdx2 sequence: sequence of the Cdx2 is shown in Figure 2. The open reading frame from our nucleic acid sequence predicts that Cdx2 contains 311 amino acids. The Cdx2 predicted protein contains the initiating methionine codon preceded by the consensus translational start site. The 3′ region consisted of a TGA stop codon at position 1004 followed by 740 nucleotides of untranslated message and a putative polyadenylation recognition signal site (TATAAA) at 1723.

Sequence alignment of Cdx1 and Cdx2

When we aligned both human Cdx1 and Cdx2 homeobox, we observed limited homology between the 2 outside the homeodomain, except for the tetrapeptide sequence EWMR, the hexapeptide PPQYPD, the amino-terminal-10 amino acids, and the glutamine-rich sequence following the homeodomain as for the mouse counterparts.

Figure 3 shows the alignment of the human and mouse Cdx1. Human Cdx1 is 85% identical to the mouse counterpart. All the 61 amino acids of the homeodomain were conserved between the human and mouse sequences. Conversely, the carboxyl-terminal regions showed only 60% identity.

Figure 4 shows alignments of the amino-acid sequence of human and mouse Cdx2, and the hamster Cdx3 homeobox. Amino-acid identity was 94% between human and mouse Cdx2, and 94% between human Cdx2 and hamster Cdx3. The homeodomain sequence was identical between human and hamster, but showed a 3-amino-acids difference with the mouse Cdx2. As for Cdx1, the carboxyl-terminal region of Cdx2 was relatively less conserved between these species. The high degree of identity between the hamster Cdx3 and the human and mouse Cdx2 (Fig. 4) supports the hypothesis that hamster Cdx3 is the homologue of human and mouse Cdx2, as suggested by Suh *et al.* (1994).

Expression of Cdx1 and Cdx2 mRNA

We studied 12 tissue samples from adenocarcinomas of the colon and the corresponding normal mucosa. Northern blots are shown in Figure 5. Methylene-blue staining of RNA was performed after blotting, to allow a rough estimate of the amount of RNA loaded onto the gel and as control of RNA integrity (data not shown). A single RNA species of about 2.0 kb hybridized to the Cdx1 3′ probe and a single mRNA species of about 1.9 kb hybridized to the Cdx2 3′ probe. In most tumour tissues, expression of this mRNA was concomitantly reduced or absent. Densitometric scanning of these blots indicated reduction from 3- to more than 100-fold. However, in 2 tumour samples (patients 4 and 7), levels of the Cdx1 and Cdx2 transcripts remained unchanged.

To examine the tissue distribution of Cdx1 and Cdx2 mRNA, Northern-blot assays were performed with RNA isolated from several human tissues. These results showed that both transcripts were expressed in small intestine, in colon and rectum. mRNA for Cdx1 and Cdx2 was not detectable in the other tissues, including thymus, spleen, adrenal gland, gall bladder, stomach, pancreas, liver and skin (Figs. 5 and 6). This pattern of expression is similar to that reported for mouse Cdx1 and Cdx2 (Duprey *et al.*, 1988; Hu *et al.*, 1993; Suh *et al.*, 1994; James and Kazenwadel, 1991) and rat Cdx1 (Freund *et al.*, 1992).

Finally, we also investigated the expression of Cdx1 and Cdx2 mRNA in 5 cell lines derived from colorectal carcinoma. As shown in Figure 6, only the Cdx1 transcript was detected in LS174T cells, whereas only Cdx2 mRNA was detected in SW480, TC7 and PF11 cells, and none of them in the HT29 cell line.

DISCUSSION

The transformation of a cell and the acquisition of the invasive and metastatic phenotypes result from the activation of complex cellular processes rather than from the effect of a single gene product. It is however likely that the coordination of the multiple genes involved in malignancy is under the control of a few genes that act as master genes. These probably code for transcription factors that control the genetic program allowing tumour invasion and metastasis. Hence, identification of those transcription factors is crucial for understanding the molecular bases of neoplasia. Homeobox genes are a family of transcription factors containing a 183-bp highly conserved nucleotide sequence coding for a 61-amino-acid domain that binds specifically to DNA. First discovered in *Drosophila* as genes controlling segmentation and segment identity (Gehring, 1990), homeobox genes have since been identified in many other species including nematodes, frog, mouse and human. Little is known on the specific function of each of these genes, although alterations in expression of several homeobox genes have been reported in a variety of malignant tissues (De Vita *et al.*, 1993; Friedmann *et al.*, 1994; Tiberio *et al.*, 1994). Qin *et al.* (1994) have demonstrated that induction of malignant transformation can be mediated by homeobox genes, through selective transcriptional repression. Thus, there is growing evidence that homeobox genes play a key role in development and differentiation and very probably in cancer.

The Cdx1 and Cdx2 homeodomain-containing genes are of particular interest in colorectal cancer, since intestine is the only organ that contains detectable levels of the corresponding transcripts (Fig. 6). This very restricted expression pattern is unusual for homeobox genes, and further indicates an important role for the protein products of these 2 genes in establishing intestinal cell identity. Moreover, in contrast to other homeobox genes expressed in intestinal cells deriving from the mesoderm, Cdx1 (Duprey *et al.*, 1988; Hu *et al.*, 1993) and Cdx2 (James *et al.*, 1994; Suh *et al.*, 1994) are expressed in endoderm-derived tissues. It is noteworthy that Cdx2 protein expression is not confined to a particular cell lineage (James *et al.*, 1994), suggesting that it may be responsible

```
    aggtgagcagtcgctggtcgtcggggcggccggcagcggcggctccagggcccagcatgcgcgggggaccccgcggc   77

caccATG TAT GTG GGC TAT GTG CTG GAC AAG GAT TCG CCC GTG TAC CCC GGC CCA GCC AGG   138
    Met Tyr Val Gly Tyr Val Leu Asp Lys Asp Ser Pro Val Tyr Pro Gly Pro Ala Arg   19

CCA GCC AGC CTC GGC CTG GGC CCG GCA AAC TAC GGC CCC CCG GCC CCG CCC CCG GCG CCC   198
Pro Ala Ser Leu Gly Leu Gly Pro Ala Asn Tyr Gly Pro Pro Ala Pro Pro Pro Ala Pro   39

CCG CAG TAC CCC GAC TTC TCC AGC TAC TCT CAC GTG GAG CCG GCC CCC GCG CCC CCG ACG   258
Pro Gln Tyr Pro Asp Phe Ser Ser Tyr Ser His Val Glu Pro Ala Pro Ala Pro Pro Thr   59

GCC TGG GGG GCG CCC TTC CCT GCG CCC AAG GAC GAC TGG GCC GCC GCC TAC GGC CCG GGC   318
Ala Trp Gly Ala Pro Phe Pro Ala Pro Lys Asp Asp Trp Ala Ala Ala Tyr Gly Pro Gly   79

CCC GCG GCC CCT GCC GCC AGC CCA GCT TCG CTG GCA TTC GGG CCC CCT CCA GAC TTT AGC   378
Pro Ala Ala Pro Ala Ala Ser Pro Ala Ser Leu Ala Phe Gly Pro Pro Pro Asp Phe Ser   99

CCG GTG CCG GCG CCC CCT GGG CCC GGC CCG GGC CTC CTG GCG CAG CCC CTC GGG GGC CCG   438
Pro Val Pro Ala Pro Pro gly Pro Gly Pro Gly Leu Leu Ala Gln Pro Leu Gly Gly Pro   119

GGC ACA CCG TCC TCG CCC GGA GCG CAG AGG CCG ACG CCC TAC GAG TGG ATG CGG CGC AGC   498
Gly Thr Pro Ser Ser Pro Gly Ala Gln Arg Pro Thr Pro Tyr Glu Trp Met Arg Arg Ser   139

GTG GCG GCC GGA GGC GGC GGT GGC AGC GGT AAG ACT CGG ACC AAG GAC AAG TAC CGC GTG   558
Val Ala Ala Gly Gly Gly Gly Gly Ser Gly Lys Thr Arg Thr Lys Asp Lys Tyr Arg Val   159

GTC TAC ACC GAC CAC CAA CGC CTG GAG CTG GAG AAG GAG TTT CAT TAC AGC CGT TAC ATC   618
Val Tyr Thr Asp His Gln Arg Leu Glu Leu Glu Lys Glu Phe His Tyr Ser Arg Tyr Ile   179

ACA ATC CGG CGG AAA TCA GAG CTG GCT GCC AAT CTG GGG CTC ACT GAA CGG CAG GTG AAG   678
Thr Ile Arg Arg Lys Ser Glu Leu Ala Ala Asn Leu Gly Leu Thr Glu Arg Gln Val Lys   199

ATC TGG TTC CAA AAC CGG CGG GCA AAG GAG CGC AAA GTG AAC AAG AAG AAA CAG CAG CAG   738
Ile Trp Phe Gln Asn Arg Arg Ala Lys Glu Arg Lys Val Asn Lys Lys Lys Gln Gln Gln   219

CAA CAG CCC CCA CAG CCG CCG ATG GCC CAC GAC ATC ACG GCC ACC CCA GCC GGG CCA TCC   798
Gln Gln Pro Pro Gln Pro Pro Met Ala His Asp Ile Thr Ala Thr Pro Ala Gly Pro Ser   239

CTG GGG GGC CTG TGT CCC AGC AAC ACC AGC CTC CTG GCC ACC TCC TCT CCA ATG CCT GTG   858
Leu Gly Gly Leu Cys Pro Ser Asn Thr Ser Leu Leu Ala Thr Ser Ser Pro Met Pro Val   259

AAA GAG GAG TTT CTG CCA TAGccccatgcccagcctgtgcgccgggggacctggggactcgggtgctgggagt   931
Lys Glu Glu Phe Leu Pro End                                                       265

gtggctcctgtgggcccaggaggtctggtccgagtctcagccctgaccttctgggacatggtggacagtcacctatcca   1010
ccctctgcatccccttggcccattgtgtgcagtaagcctgttggataaagaccttccagctcctgtgttctagacctct   1089
gggggataagggagtccagggtggatgatctcaatctcccgtgggcatctcaagccccaaatggttggggaggggcct   1168
agacaaggctccaggccccacctcctcctccatacgttcagaggtgcagctggaggcctgtgtggggaccacactgatc   1247
ctggagaaaagggatggagctgaaaaagatggaatgcttgcagagcatgacctgaggagggaggaacgtggtcaactca   1326
cacctgcctcttctgcagcctcacctctacctgcccccatcataagggcactgagcccttcccaggctggatactaagc   1405
acaaagcccatagcactgggctctgatggctgctccactgggttacagaatcacagccctcatgatcattctcagtgag   1484
ggctctggattgagagggaggccctgggaggagagaagggggcagagtcttccctaccaggtttctacacccccgccag   1563
gctgcccatcagggcccagggagcccccagaggactttattcggaccaagcagagctcacagctggacaggtgttgtat   1642
atagagtggaatctcttggatgcagcttcaagaataaattttcttctctttcaaaaaaaaaaaaaaaa             1711
```

FIGURE 1 – Sequence of human Cdx1 mRNA and deduced sequence of the encoded protein. The nucleotide sequence of the mRNA was deduced from that of the cDNA insert of clone 6G2 and from a cloned RT-PCR-amplification product (see "Material and Methods") encoding the 5′ region of the Cdx1 mRNA. The open reading frame encoding human Cdx1 is given in capital letters. Non-coding sequences are in lower-case letters, and the polyadenylation site AATAAA is underlined.

for establishing regional rather than cellular identity in intestinal epithelium. Furthermore, since its expression is tissue-specific and is present from the early embryo to the adult, it is likely that Cdx2 has a role both in establishing and in maintaining the intestinal phenotype, as suggested by Suh and Traber (1996). In addition, detection by *in situ* hybridization of the mouse Cdx1 transcript in the intestine begins at about 14 days *post coitum* (Duprey *et al.*, 1988), although primitive gut development begins much earlier, at 8 days *post coitum*. Thus, Cdx1 functions in intestinal-tract cell differentiation must occur late in the terminal differentiation process, probably during cytodifferentiation. Finally, it has been shown that the intestine-specific transcription of sucrase-isomaltase, a gene expressed exclusively in differentiated enterocytes, is dependent on the binding of Cdx2 to an evolutionary conserved promoter element in the sucrase-isomaltase gene (Suh *et al.*, 1994). Taken together, these findings suggest that members of the Cdx gene family play a fundamental role both in the establishment of the intestinal phenotype during development and in maintenance of this phenotype via transcriptional activation of intestinal genes of the differentiated phenotype.

We found that 10 out of 12 human colorectal carcinomas showed little or no Cdx1 and Cdx2 mRNA expression (Fig. 5), whereas 2 out of 12 expressed these transcripts at normal levels. We have observed an inverse relationship between Cdx1 and Cdx2 mRNA

```
   gcgcccctggcagccttcaacgtcggtccccaggcagcatggtgaggtctgctcccggaccctcgccaccATG TAC   76
                                                                           Met Tyr   2
GTG AGC TAC CTC CTG GAC AAG GAC GTG AGC ATG TAC CCT AGC TCC GTG CGC CAC TCT GGC   136
Val Ser Tyr Leu Leu Asp Lys Asp Val Ser Met Tyr Pro Ser Ser Val Arg His Ser Gly   22

GGC CTC AAC CTG GCG CCG CAG AAC TTC GTC AGC CCC CCG CAG TAC CCG GAC TAC GGC GGT   196
Gly Leu Asn Leu Ala Pro Gln Asn Phe Val Ser Pro Pro Gln Tyr Pro Asp Tyr Gly Gly   42

TAC CAC GTG GCG GCC GCA GCT GCA GCG CAG AAC TTG GAC AGC GCG CAG TCC CCG GGG CCA   256
Tyr His Val Ala Ala Ala Ala Ala Ala Gln Asn Leu Asp Ser Ala Gln Ser Pro Gly Pro   62

TCC TGG CCG GCA GCG TAT GGC GCC CCA CTC CGG GAG GAC TGG AAT GGC TAC GCG CCC GGA   316
Ser Trp Pro Ala Ala Tyr Gly Ala Pro Leu Arg Glu Asp Trp Asn Gly Tyr Ala Pro Gly   82

GGC GCG GCC GCC GCC AAC GCC GTG GCT CAC GCG CTC AAC GGT GGC TCC CCG GCC GCA GCC   376
Gly Ala Ala Ala Ala Asn Ala Val Ala His Ala Leu Asn Gly Gly Ser Pro Ala Ala Ala   102

ATG GGC TAC AGC AGC CCC GCA GAC TAC CAT CCG CAC CAC CAC CCG CAT CAC CAC CCG CAC   436
Met Gly Tyr Ser Ser Pro Ala Asp Tyr His Pro His His His Pro His His His Pro His   122

CAC CCG GCC GCC GCG CCT TCC TGC GCT TCT GGG CTG CTG CAA ACG CTC AAC CCC GGC CCT   496
His Pro Ala Ala Ala Pro Ser Cys Ala Ser Gly Leu Leu Gln Thr Leu Asn Pro Gly Pro   142

CCT GGG CCC GCC GCC ACC GCT GCC GCC GAG CAG CTG TCT CCC GGC GGC CAG CGG CGG AAC   556
Pro Gly Pro Ala Ala Thr Ala Ala Ala Glu Gln Leu Ser Pro Gly Gly Gln Arg Arg Asn   162

CTG TGC GAG TGG ATG CGG AAG CCG GCG CAG CAG TCC CTC GGC AGC CAA GTG AAA ACC AGG   616
Leu Cys Glu Trp Met Arg Lys Pro Ala Gln Gln Ser Leu Gly Ser Gln Val Lys Thr Arg   182

ACG AAA GAC AAA TAT CGA GTG GTG TAC ACG GAC CAC CAG CGG CTG GAG CTG GAG AAG GAG   676
Thr Lys Asp Lys Tyr Arg Val Val Tyr Thr Asp His Gln Arg Leu Glu Leu Glu Lys Glu   202

TTT CAC TAC AGT CGC TAC ATC ACC ATC CGG AGG AAA GCC GAG CTA GCC GCC ACG CTG GGG   736
Phe His Tyr Ser Arg Tyr Ile Thr Ile Arg Arg Lys Ala Glu Leu Ala Ala Thr Leu Gly   222

CTC TCT GAG AGG CAG GTT AAA ATC TGG TTT CAG AAC CGC AGA GCA AAG GAG AGG AAA ATC   796
Leu Ser Glu Arg Gln Val Lys Ile Trp Phe Gln Asn Arg Arg Ala Lys Glu Arg Lys Ile   242

AAC AAG AAG AAG TTG CAG CAG CAA CAG CAG CAG CAG CCA CCA CAG CCG CCT CCG CCG CCA   856
Asn Lys Lys Lys Leu Gln Gln Gln Gln Gln Gln Gln Pro Pro Gln Pro Pro Pro Pro Pro   262

CCA CAG CCT CCC CAG CCT CAG CCA GGT CCT CTG AGA AGT GTC CCA GAG CCC TTG AGT CCG   916
Pro Gln Pro Pro Gln Pro Gln Pro Gly Pro Leu Arg Ser Val Pro Glu Pro Leu Ser Pro   282

GTG TCT TCC CTG CAA GCC TCA GTG TCT GGC TCT GTC CCT GGG GTT CTG GGG CCA ACT GGG   976
Val Ser Ser Leu Gln Ala Ser Val Ser Gly Ser Val Pro Gly Val Leu Gly Pro Thr Gly   302

GGG GTG CTA AAC CCC ACC GTC ACC CAG TGAcccaccggggtctgcagcggcagagcaattccaggctgag   1046
Gly Val Leu Asn Pro Thr Val Thr Gln End                                           311

ccatgaggagcgtggactctgctagactcctcaggagagacccctcccctcccacccacagccatagacctacagacct   1125
ggctctcagaggaaaaatgggagccaggagtaagacaagtgggatttggggcctcaagaaatatactctcccagatttt   1204
tactttttccatctggcttttctgccactgaggagacagaaagcctccgctgggcttcattccggactggcagaagca   1283
ttgcctggactgaccacaccaaccagcttcatctatccgactcttctcttcctagatctgcaggctgcacctctggcta   1362
gagccgaggggagagagggactcaagggaaaggcaagcttgaggccaagatggctgctgcctgctcatggccctcggag   1441
gtccagctgggcctcctgcctccgggcagcaaggtttacactgcggaacgcaaaggcagctaagatagaaagctggact   1520
gaccaaagactgcagaacccccaggtggccctgcgtctttttctcttccctttcccagaccaggaaaggcttggctgg   1599
tgtatgcacagggtgtggtatgagggggtggttattggactccaggcctgaccaggggcccgaacaggacttgttaga   1678
gagcctgtcaccagagcttctctgggctgaatgtatgtcagtgctataaatgccagagccaacctgaaaaaaa          1751
```

FIGURE 2 – Sequence of human Cdx2 mRNA and deduced sequence of the encoded protein. The nucleotide sequence of the mRNA was deduced from that of the cDNA insert of clone 32D6, clone λcdx2a and from a cloned RT-PCR-amplification product (see "Material and Methods") encoding the 5′ region of the Cdx2 mRNA. The open reading frame encoding human Cdx2 is given in capital letters. Non-coding sequences are in lower-case letters, and the putative polyadenylation site TATAAA is underlined.

levels and the severity of dysplasia and the clinical outcome (data not shown). Although the number of cases analyzed in this study is too small to draw any definitive conclusion, this suggests that decreased Cdx1 and/or Cdx2 expression is associated with progression to the more advanced stages of colorectal tumorigenesis. However, it remains possible that alteration of Cdx gene expression is a consequence of the dysregulation that accompanies tumour progression. The target genes of the transcription factors encoded by the homeotic genes are unknown, but some reports indicate that they may regulate the expression of adhesion molecules on the membrane (Edelman and Jones, 1992; Walsh and Doherty, 1993) and oncogenes (Qin *et al.*, 1994). Abnormal expression of these genes could therefore allow spreading of cells to aberrant locations. A more clinically oriented analysis involving a larger number of

```
HumCdx1   M Y V G Y V L D K D S P V Y P G P A R P A S L G L G P A N Y G P P A P P P A P P Q Y P D F S S Y S H    50
MiceCdx1  M Y V G Y V L D K D S P V Y P G P A R P S S L G L G P P T Y A P P G P A P A P P Q Y P D F A G Y T H    50

HumCdx1   V E P A P A P P T A W G A P F P A P K D D W A A A Y G P G P A A P A A S P A S L A F G P P P D F S P   100
MiceCdx1  V E P A P A P P P T W A A P F P A P K D D W A A A Y G P G P T A S A A S P A P L A F G P P P D F S P   100

HumCdx1   V P A P P G P G P G L L A Q P L G G P G T P S S P G A Q R P T P Y E W M R R S V A A G G G G G S G K   150
MiceCdx1  V P A P P G P G P G I L A Q S L G A P G A P S S P G A P R R T P Y E W M R R S V A A A G G G G S G K   150

HumCdx1   T R T K D K Y R V V Y T D H Q R L E L E K E F H Y S R Y I T I R R K S E L A A N L G L T E R Q V K I   200
MiceCdx1  T R T K D K Y R V V Y T D H Q R L E L E K E F H Y S R Y I T I R R K S E L A A N L G L T E R Q V K I   200

HumCdx1   W F Q N R R A K E R K V N K K K Q Q Q Q Q P P Q P P M A H - - - D I T A T P A G P S L G G L C P S N   247
MiceCdx1  W F Q N R R A K E R K V N K K K Q Q Q Q Q P L P P T Q L P L P L D G T P T P S G P P L G S L C P T N   250

HumCdx1   T S L L A T S S P M P V K E E F L P                                                                 265
MiceCdx1  A G L L G T P S P V P V K E E F L P                                                                 268
```

FIGURE 3 – Amino-acid-sequence comparison between human (HumCdx1) and mice (MiceCdx1) Cdx1. Boxed areas correspond to amino-acid identities. The homeobox domain is shown in bold letters.

```
HumCdx2   MYVSYLLDKDVSMYPSSVRHSGGLNLAPQNFVSPPQYPDYGGYHVAAAAA   50
MiceCdx2  MYVSYLLDKDVSMYPSSVRHSGGLNLAPQNFVSPPQYPDYGGYHVAAAAA   50
HamCdx3   MYVSYLLDKDVSMYPSSVRHSGGLNLAPQNFVSPPQYPDYGGYHVAAAAA   50

HumCdx2   A-QNLDSAQSPGPSWPAAYGAPLREDWNGYAPGGAAAANAVAHALNGGSP   99
MiceCdx2  ATANLDSAQSPGPSWPTAYGAPLREDWNGYAPGGAAAANAVAHGLNGGSP  100
HamCdx3   AAANLDSAQSPGPSWSTAYGAPLREDWNGYPPGGAAAANAVAHGLNGGSP  100

HumCdx2   AAAMGYSSPADYHPHHHPHHHPHHPAAAPSCASGLLQTLNPGPPGPAATA  149
MiceCdx2  AAAMGYSSPAEYHAHHHPHHHPHHPAASPSCASGLLQTLNLGPPGPAATA  150
HamCdx3   AAAMGYSSPADYHAHHHPHHHPHHPAAAPSCASGLLQTLNPGPPGPAATG  150

HumCdx2   AAEQLSPGGQRRNLCEWMRKPAQQSLGSQVKTRTKDKYRVVYTDHQRLEL  199
MiceCdx2  AAEQLSPSGQRRNLCEWMRKPAQQSLGSQVKTRTKDKYRVVYTDHQRLEL  200
HamCdx3   AAEQLSPSGQRRNLCEWMRKPAQPSLGSQVKTRTKDKYRVVYTDHQRLEL  200

HumCdx2   EKEFHYSRYITIRRKAELAATLGLSERQVKIWFQNRRAKERKINKKKL-Q  248
MiceCdx2  EKEFHFSRYITIRRKSELAATLGLSERQVKIWFQNRRAKERKIKKK--QQ  248
HamCdx3   EKEFHYSRYITIRRKAELAATLGLSERQVKIWFQNRRAKERKINKKKLQQ  250

HumCdx2   QQQQQQPPQPPPPPPQPPQPQPGPLRSVPEPLSPVSSLQASVSGSVPGVL  298
MiceCdx2  QQQQQQQQQPPQPPPQPSQPQPGALRSVPEPLSPVTSLQGSVPGSVPGVL  298
HamCdx3   QQQQQQQQQLASPPPQPSQPQPGSLRSVPEPLSPVSSLQGSVPGSVPGVL  300

HumCdx2   GPTGGVLNPTVTQ                                       311
MiceCdx2  GPAGGVLNSTVTQ                                       311
HamCdx3   GPAGGVLNPTVTQ                                       313
```

FIGURE 4 – Amino-acid-sequence comparison between human (HumCdx2), mice (MiceCdx2) Cdx2 and hamster Cdx3 (HamCdx3). Boxed areas correspond to amino-acid identities. The homeobox domain is shown given in bold letters.

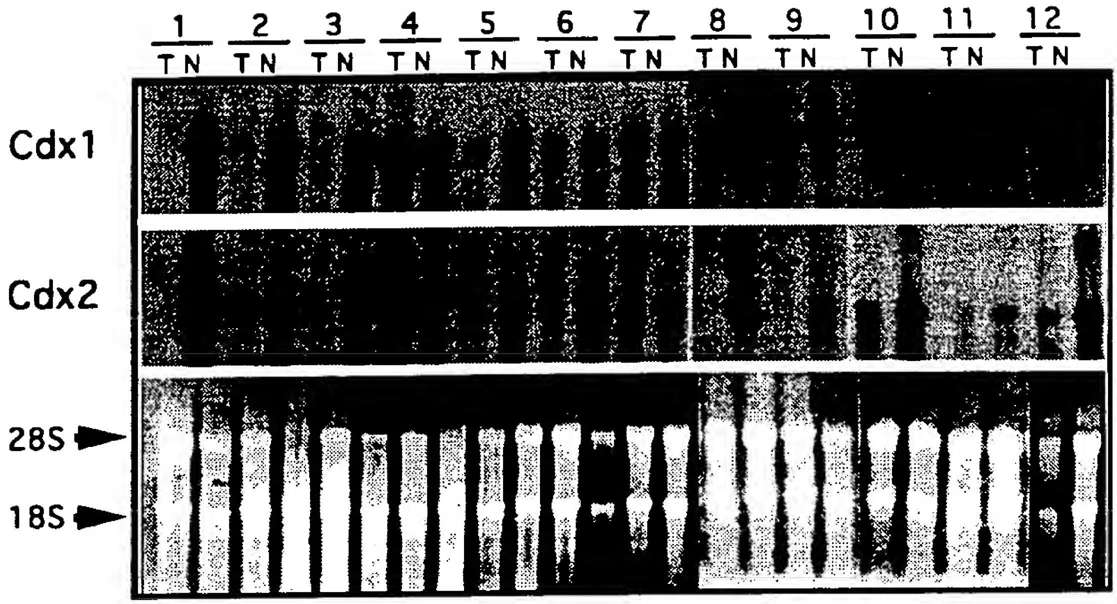

FIGURE 5 - Cdx1 and Cdx2 mRNA expression in colorectal carcinoma. Northern-blot analysis using 20 µg total RNA from colon tumors (T) and matched normal tissues (N) surgically removed from patients. Patients (n = 12) are identified by figures on top of the panels. The quality of RNA preparations was verified by running 10 µg on agarose gel and controlling the relative amounts of 18s and 28s ribosomal RNA. Quality and quantity of RNA on the filters were controlled by methylene-blue coloration (data not shown). cDNA inserts from clones 6G2 (Cdx1) and 32D6 (Cdx2) labeled with α[32P]dCTP by random priming were used as probes. The autoradiographs were exposed at −80°C for 3 days.

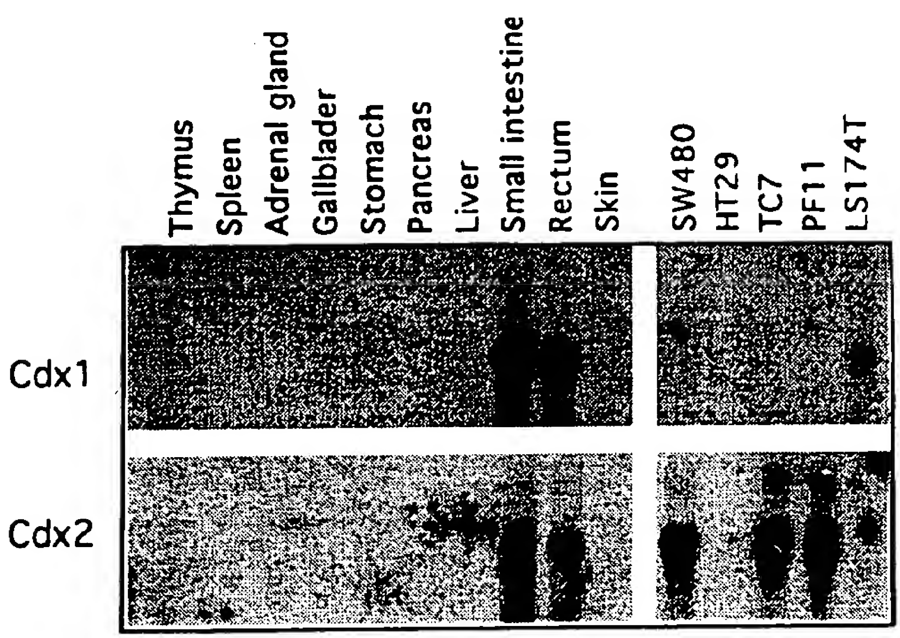

FIGURE 6 - Tissue specificity of human Cdx1 and Cdx2 mRNA expression. Total RNA (20 µg) was isolated from various human tissues and 5 colon-carcinoma-derived cell lines as described in "Material and Methods", denatured, and electrophoresed in a 1% agarose-formaldehyde gel. The quality of RNA preparations was verified by running 10 µg on agarose gel and controlling the relative amounts of 18S and 28S ribosomal RNA. After capillary pressure transfer to nylon membranes, the blots were hybridized with cDNA inserts from clones 6G2 (Cdx1) and 32D6 (Cdx2) labeled with α[32P]dCTP by random priming. Quality and quantity of RNA on the filters were controlled by methylene-blue coloration (data not shown). The autoradiographs were exposed at −80°C for 3 days.

cases is needed to characterize the relation between the pattern of expression of Cdx genes and colon-cancer progression.

As judged from Figure 5, Cdx1 and Cdx2 mRNA expression was always coordinately modulated during colorectal carcinogenesis, suggesting a common regulatory mechanism for expression of the 2 genes. However, although Cdx1 and Cdx2 genes were present into the genome of the 5 colorectal-cancer-derived cell lines (data not shown), in 4 of them expression of the 2 transcripts was not coordinately regulated (LS174T, PF11, TC7 and SW480 cells, Fig. 6). Hence Cdx1 mRNA and Cdx2 mRNA are modulated through independent pathways. Yet, the 2 pathways must share some steps, those affected during colorectal carcinogenesis and responsible for the observed coordinated down-regulation.

A number of studies have compared homeobox gene expression in tumours and their corresponding normal tissues. Friedmann *et al.* (1994) reported that HOXC-6 was expressed in the normal mammary gland of mice, at low level in pre-cancerous tissue and not expressed in cancer. Tiberio *et al.* (1994) showed that down-regulation of HOX genes plays a role in small-cell-lung-cancer progression, possibly through alteration in their tumour-suppression activity. Whereas most homeobox genes were down-regulated during carcinogenesis, other were unmodified or up-regulated (reviewed in Barba *et al.*, 1993). De Vita *et al.* (1993) have analyzed HOX-gene expression in normal colon and in colorectal carcinoma. Although the expression of some HOX genes is identical in normal and neoplastic colon, others such as HOX2C and HOX4B display alterations in the size of their transcripts, due to alternative splicing. Unfortunately, these authors did not perform quantitative analysis of these homeobox mRNAs. Quantitative and qualitative changes in the expression of most, but not all, homeobox genes suggest that they might participate in carcinogenesis as suppressor genes, if down-regulated, or as oncogenes if up-regulated. As down-regulated transcriptional factors, Cdx1 and Cdx2 homeobox are 2 potential suppressor genes. Whether or not Cdx1 or Cdx2 are directly involved in the process of intestinal carcinogenesis remains to be elucidated.

While writing this text, we learned that Bonner *et al.* (1995) have reported partial sequence of the human Cdx1 gene (Genbank accesion numbers U15212 and U16360). Their sequence overlaps with our sequence between positions 1 and 1529. The translated region shows perfect identity; however, within the 3′ untranslated region, one G is inserted at position 1526 and the C (position 1528) is replaced by A in our sequence.

The nucleotide sequences reported in this paper have been submitted to the GenBank Data Bank with accession numbers U51095 (Cdx1) and U51096 (Cdx2).

ACKNOWLEDGEMENTS

G.V.M. is a fellow of the Fondation pour la Recherche Médicale; J-M.F. was supported by a post-doctoral fellowship from the Association pour la Recherche contre le Cancer, and N.D. was supported by the Institut National de la Santé et de la Recherche Médicale (Poste Vert). This work was supported by grants from the Groupement de Recherches et d'Études sur les Génomes (GREG), the Ligue Nationale Contre le Cancer (Comité des Bouches du Rhône), the Association pour la Recherche sur le Cancer (ARC) and the Fédération Nationale des Centres de Lutte Contre le Cancer (FLNCC).

REFERENCES

BARBA, P., MAGLI, M.C., TIBERIO, C. and CILLO, C., HOX gene expression in human cancers. *Advanc. exp. Med. Biol.*, **348**, 45–57 (1993).

BARNARD, G.F., STANIUNAS, R.J., BAO, S., MAFUNE, K., STEELE, G.D., GOLLAN, J.L. and CHEN, L.B., Increased expression of human ribosomal phosphoprotein P0 messenger RNA in hepatocellular carcinoma and colon carcinoma. *Cancer Res.*, **52**, 3067–3072 (1992*a*).

BARNARD, G.F., STANIUNAS, R.J., MORI, M., PUDER, M., JESSUP, M.J., STEELE, G.D. and CHEN, L.B., Gastric and hepatocellular carcinomas do not over-express the same ribosomal protein messenger RNA as colonic carcinoma. *Cancer Res.*, **53**, 4048–4052 (1992*b*).

BARTSCH, R.A., JOANNOU, C., TALBOT, I.C. and BAILEY, D.S., Transcription factors and the control of *Drosophila* development. *Brit. J. Cancer*, **54**, 791–798 (1986).

BONNER, C.A., LOFTUS, S.K. and WASMUTH, J.J., Isolation, characterization and precise physical localization of human CDX1, a caudal-type homeobox gene. *Genomics*, **28**, 206–211 (1995).

CHANTRET, I., BARBAT, A., DUSSAULX, E., BRATTAIN, M.G. and ZWEIBAUM, A., Epithelial polarity, villin expression and enterocytic differentiation of cultured human colon-carcinoma cells: a survey of twenty cell lines. *Cancer Res.*, **48**, 1936–1942 (1988).

CHIRGWIN, J.M., PRZYBYLA, A., MACDONALD, R.J. and RUTTER, W.J., Isolation of biologically active ribonucleic acif from sources enriched in ribonuclease. *Biochemistry*, **24**, 5294–5299 (1979).

CHOMCZYNSKI, P. and SACCHI, N., Single-step method of RNA isolation by acid guanidinium-thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, **162**, 156–159 (1987).

DE VITA, G., BARBA, P., ODARTCHENKO, N., GIVEL, J.C., FRESCHI, G., BUCCIARELLI, G., MAGLI, M.C., BONCINELLI, E. and CILLO, C., Expression of homeobox-containing genes in primary and metastatic colorectal cancer. *Europ. J. Cancer*, **29**, 887–893 (1993).

DENIS, M.G., CHADENEAU, C., LECABELLEC, M.T., LEMOULLAC, B., LEMEVEL, B., MEFLAH, K. and LUSTENBERGER, P., Over-expression of the S13 ribosomal protein in actively growing cells. *Int. J. Cancer*, **55**, 275–280 (1993).

DON, R.H., COX, P.T., WANWRIGHT, B.J., BAKER, K. and MATTICK, J.S., "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.*, **19**, 4008 (1991).

DUPREY, P., CHOWDHURY, K., DRESSLER, G R., BALLING, R., SIMON, D., GUENET, J.L. and GRUSS, P., A mouse gene homologous to the *Drosophila* gene caudal is expressed in epithelial cells from the embryonic intestine. *Genes Dev.*, **2**, 1647–1654 (1988).

DUSETTI, N., FRIGERIO, J.M., KEIM, V., DAGORN, J.C. and IOVANNA, J.L., Structural organization of the gene encoding the rat pancreatitis-associated protein. Analysis of its evolutionary history reveals an ancient divergence from the other carbohydrate-recognition-domain-containing genes. *J. biol. Chem.*, **268**, 14470–14475 (1993).

EDELMAN, G.M. and JONES, F.S., Cytotactin: a morpho-regulatory molecule and a target for regulation by homeobox-gene products. *Trends biochem. Sci.*, **17**, 228–232 (1992).

FREUND, J.N., BOUKAMEL, R. and BENAZZOUZ, A., Gradient expression of Cdx along the rat intestine throughout post-natal development. *FEBS Lett.*, **314**, 163–166 (1992).

FRIEDMANN, Y., DANIEL, C.A., STRICKLAND, P. and DANIEL, C.W., Hox genes in normal and neoplastic mouse mammary gland. *Cancer Res.*, **54**, 5981–5985 (1994).

FRIGERIO, J.M., BERTHEZENE, P., GARRIDO, P., ORTIZ, E., BERTHELLEMY, S., VASSEUR, S., SASTRE, B., SELEZNIEFF, I., DAGORN, J.C. and IOVANNA, J.L., Analysis of 2166 clones from a human colorectal-cancer cDNA library by partial sequencing. *Hum. mol. Genet.*, **4**, 37–43 (1995).

GEHRING, W.J., MULLER, M., AFFOLTER, M., PERCIVAL-SMITH, A., BILLETER, M., QIAN, Y.Q., OTTING, G. and WUTHRICH, K., The structure of the homeodomain and its functional implications. *Trends Genet.*, **6**, 323–329 (1990).

GERMAN, M.S., WANG, J., CHADWICK, R.B. and RUTTER, W.J., Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: building a functional insulin mini-enhancer complex. *Genes Dev.*, **6**, 2165–2176 (1992).

HU, Y., KAZENWADEL, J. and JAMES, R., Isolation and characterization of the murine homeobox gene Cdx-1. Regulation of expression in intestinal epithelial cells. *J. biol. Chem.*, **268**, 27214–27225 (1993).

JAMES, R., ERLER, T. and KAZENWADEL, J., Structure of the murine homeobox gene cdx-2. Expression in embryonic and adult intestinal epithelium. *J. biol. Chem.*, **269**, 15229–15237 (1994).

JAMES, R. and KAZENWADEL, J., Homeobox gene expression in the intestinal epithelium of adult mice. *J. biol. Chem.*, **266**, 3246–3251 (1991).

KONDOH, N., SCHWEINFEST, C.L., HENDERSON, K.W. and PAPAS, T.S.,

Differential expression of S19 ribosomal protein, laminin-binding protein, and human lymphocyte antigen class-I messenger RNA associated with colon-carcinoma progression and differentiation. *Cancer Res.*, **52**, 791–796 (1992).

NGUYEN, C., ROCHA, D., GRANJEAUD, S., BALDIT, M., BERNARD, K., NAQUET, P. and JORDAN, B., Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. *Genomics*, **29**, 207–216 (1995).

NIZETIC, D., DRMANAC, R. and LEHRACH, H., An improved bacterial colony lysis procedure enables direct DNA hybridisation using short (10, 11 bases) oligonucleotides to cosmids. *Nucleic Acids Res.*, **19**, 182 (1991).

QIN, X.F., LUO, Y., SUH, H., WAYNE, J., MISULOVIN, Z., ROEDER, R.G. and NUSSENZWEIG, M.C., Transformation by homeobox genes can be mediated by selective transcriptional repression. *EMBO J.*, **13**, 5967–5976 (1994).

SAMBROOK, J., FRITSCH, E.F. and MANIATIS, T., *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory, NY (1989).

SCHWEINFEST, C.W., HENDERSON, K.W., SUSTER, S., KONDOH, N. and PAPAS, T.S., Identification of a colon-mucosa gene that is down-regulated in colon adenomas and adenocarcinomas. *Proc. nat. Acad. Sci. (Wash.)*, 90, 4166–4170 (1993).

SUH, E., CHEN, L., TAYLOR, J. and TRABER, P.G., A homeodomain protein related to caudal regulates intestine-specific gene transcription. *Mol. cell. Biol.*, **14**, 7340–7351 (1994).

SUH, E. and TRABER, P.G., An intestine-specific homeobox gene regulates proliferation and differentiation. *Mol. cell. Biol.*, **16**, 619–625 (1996).

TIBERIO, C., BARBA, P., MAGLI, M.C., ARVELO, F., LE CHEVALIER, T., POUPON, M.F. and CILLO, C., HOX gene expression in human small-cell lung cancers xenografted into nude mice. *Int. J. Cancer*, **58**, 608–615 (1994).

WALSH, F.S. and DOHERTY, P., Factors regulating the expression and function of calcium-independent cell-adhesion molecules. *Curr. Opin. Cell Biol.*, **5**, 791–796 (1993).

YOW, H., WONG, J.M., CHEN, H.S., LEE, C., DAVIS, S., STEELE, G.D. and CHEN, L.B., Increased mRNA expression of a laminin-binding protein in human colon carcinoma: complete sequence of a full-length cDNA encoding the protein. *Proc. nat. Acad. Sci. (Wash.)*, **85**, 6394–6398 (1988).